

In sieben Schritten zum Datenmodell ...? Das Projekt ASCH

Susanne Al-Eryani
27.02.2016



Projektdaten



- **Titel:** Entwicklung von interoperablen Standards für die Kontextualisierung heterogener Objekte am Beispiel der Provenienz Asch
- **Projektförderung:** Deutsche Forschungsgemeinschaft
- **Laufzeit:** 01.09.2014 – 31.08.2017
- **Projektleitung:**
 - Niedersächsische Staats- und Universitätsbibliothek Göttingen
 - Institut für Ethnologie der Universität Göttingen
- **Kontakt:**
 - **Projektkoordination:** Stefanie Rühle (sruehle(a)sub.uni-goettingen.de)
 - Susanne Al-Eryani (al-eryani(a)sub.uni-goettingen.de)
 - Gudrun Bucher (gudrun.bucher(a)sub.uni-goettingen.de)
 - Jürgen Dönitz (doenitz(a)sub.uni-goettingen.de)



<https://goo.gl/maps/hChbV9dP4zJ2>



Kunstsammlung



SUB



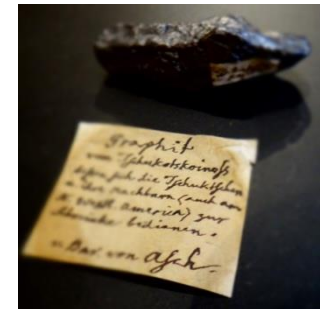
Münzsammlung



Ethnologische Sammlung



Königlich Academisches Museum, 1773



Historische Sammlungen des Geowissenschaftlichen Zentrums



Schädelsammlung

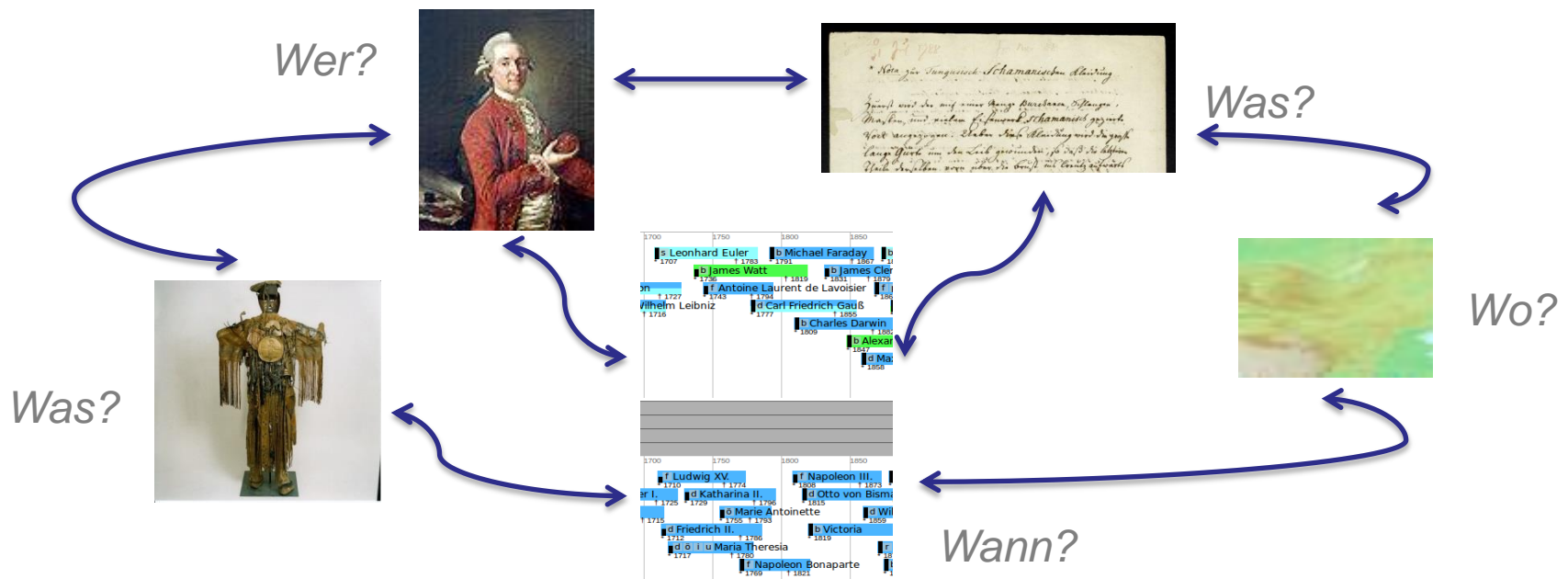


Zoologische Sammlung

Ziel des Projekts



- Entwicklung eines Metadatenmodells für die Beschreibung
 - von Provenienzinformationen (Evidenzen) zu Objekten aus unterschiedlichen Sammlungen
 - von unterschiedlichen Objekttypen
 - von Beziehungen zwischen Fakten / Entitäten (Objekten, Personen, Orten, Ereignissen, ...)

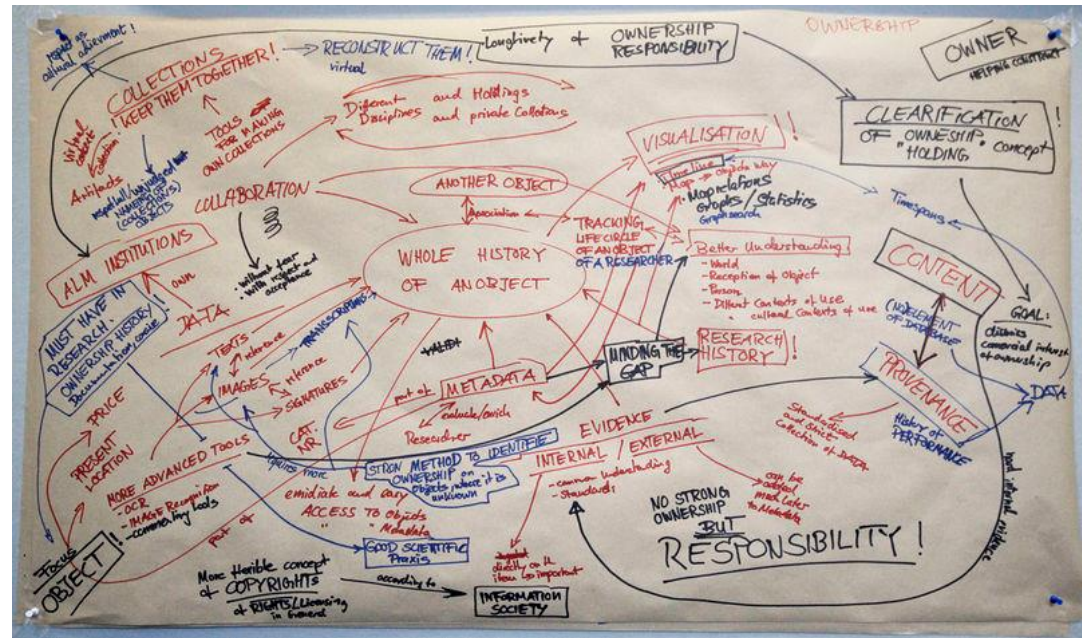


Schritt 1: Erhebung



- Workshop mit Experten aus Geistes- und Naturwissenschaften
- einrichtungsübergreifend
 - Museen
 - Bibliotheken
 - Archive
- fächerübergreifend
 - Anatomie / Humanbiologie
 - Archäologie
 - Bibliothekswissenschaft
 - Botanik
 - Ethnologie
 - Geologie / Mineralogie
 - Informatik
 - Kunstgeschichte
 - Musikwissenschaft
 - Zoologie

Interviews

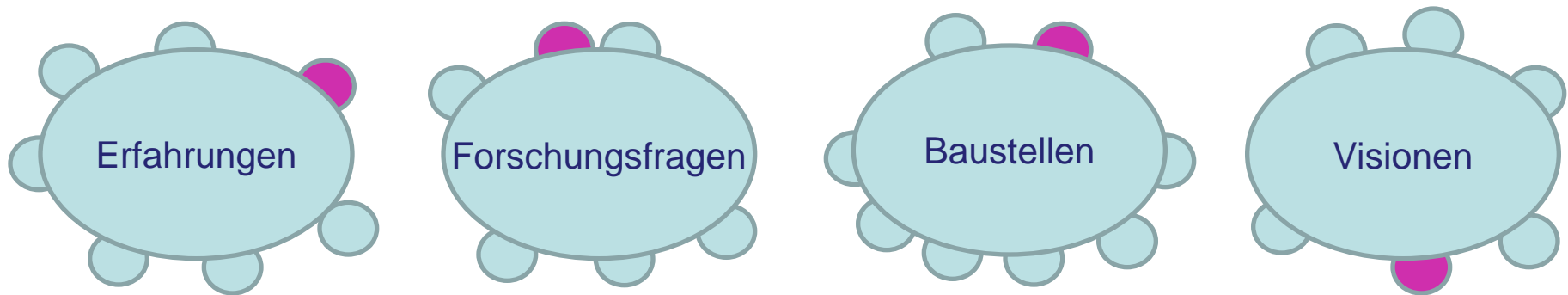


Change of ownership / Chain of custody – Visions
http://asch.wiki.gwdg.de/index.php/Workshop_2015

Workshop und Interviews: Themenschwerpunkte



- Umgang mit Provenienzinformatoren zu Sammlungsobjekten
- Datenmanagement
 - Welche Informationen liegen zu den Objekten vor (Katalog, Etikett, Archivmaterial, Beschriftungen auf Kisten, Publikationen...)?
 - Inwieweit sind diese Informationen zusammengetragen / zugänglich?
 - Welche Datenbank wird genutzt?
 - Welche Daten befinden sich in der Datenbank?



➤ verschiedene Datenlieferanten – verschiedene institutionale / fachliche Profile

- ➡ unterschiedliche Datenformate
- ➡ unterschiedliche Vokabulare
- ➡ unterschiedliche Anforderungen

Schritt 2: Auswertung

a) Formulierung von Case Studies und Scenarios



A	B	C	D
ID	Case Study	Scenario ID	Scenario
CS 1	For his work on Russian-German relationships in the 18 th century Hans needs to know more about Baron Georg T. von Asch and the items he sent to the University of Göttingen. During his search he learns more about von Asch's studies in Göttingen and that he sent items from different parts of Asia and Siberia to the University later on, which today are dispersed over several academic collections. Next to information about von Asch he finds a list of terms, categorizing the different items related to von Asch (Geology, Zoology, Ethnography, Literature, Letters, etc.) and information about the collections, holding these items today. For every item he gets a detailed description and a digital representation.	Scenario 6	A user needs biographical information about life and work of an agent.
		Scenario 7	A user is searching for all items related to a person.
		Scenario 26	A user wants to know more about places related to an item.
		Scenario 43	A user found information about an item and wants to know how he/she may get access to the item.
		Scenario 14	Browsing the web a user has found information about a collection. He/she wants to know what sort of items he/she will find at the collection.
		Scenario 15	A user wants to qualify his/her search using values from a controlled vocabulary (e.g. a taxonomy or authority file).
CS 2	Greta is working on an article about fluorit occurrence at the Yenisei. She is living in Hannover and hopes to get some samples at a university nearby. She is searching for "fluorite" and gets a list of items named as "fluorite" or "fluorspar". The list is ranked by the proximity of the holding institutions to her location. She specifies her search by using the geographic coordinates of the area she is working about. She gets a shorter list and chooses the first hit. There she finds a detailed description of fluorit at the University of Göttingen and a link to the collection the fluorit belongs to and a link to a contact address for further information.	Scenario 13	Searching for an item a user uses a controlled vocabulary to get all items of the same class.
		Scenario 17	A user is identifying the item he/she needs by a digital representation of the item.
		Scenario 12	A user is searching for a resource known by different names. He/she uses one of the names for his/her search and gets all hits where one of the names is been used.
		Scenario 45	A user is searching for an item in an institution nearby. He/she gets a list of all institutions storing such items and wants to select the nearest to his/her location.
		Scenario 27	A user is searching for items that have been collected in a certain area.



b) Formulierung von Use Cases

- UC 1 Information über Ressourcen
- UC 2 Identifikation von Ressourcen
- UC 3 Information über die Lebensgeschichte von Ressourcen
- UC 4 Wandel in Gebrauch und Rezeption von Ressourcen
- UC 5 Untermauerung der Informationen durch Evidenzen
- UC 6 Provenienz von Aussagen
- **UC 7 Zugang zu Ressourcen**
- UC 8 Wiederverwendbarkeit der Daten

(http://asch.wiki.gwdg.de/index.php/Use_Cases)



c) Verknüpfung von Use Cases und Scenarios

Use Case	ID	Scenario	Actor	Goal
UC 6 Provenance of statements	Scenario 42	An editor describes the origins of his/her statements.	editor	Specify a statement by statements affirming this relation.
Scenarios concerning the usability of resources				
UC 7 Access	Scenario 43	A user found information about an item and wants to know how he/she may get access to the item.	user	Browse from item description to the description of the terms of use Browse from item description to the description of the repository/place
UC 7 Access	Scenario 44	A user wants to contact the agents who may help him/her to use items of a collection.	user	Browse from item description to the description of the repository/place Find contact information at the description of the repository/place
UC 7 Access	Scenario 45	A user is searching for an item in an institution nearby. He/she gets a list of all institutions storing such items and wants to select the nearest to his/her location.	user	Find a resource near to your location.
UC 7 Access	Scenario 46	A user is searching for a collection nearby.	user	Find a collection near to your location.

Schritt 3: Anforderungsanalyse



➤ **Bedürfnisse des End-Nutzers**

- Nutzungskontext: Für welchen Bereich wird das Modell entwickelt?
- Zielgruppe: Wer soll die Daten nutzen?
- Sprache?

➤ **Anforderungen an Metadaten**

- Welche Eigenschaften und welche Beziehungen von Metadaten sollen im Modell abgebildet werden?

➤ **Anforderungen an System**

- Bereitstellung / Verfügbarkeit: Wie sollen die Daten später abrufbar sein?
- Funktionen: Wie sollen die Daten auffindbar sein? Wie sind die Daten dargestellt?



Formulierung der Requirements

ID	Scenario	Actor	Goal	Requirement ID	Requirement
Scenario 42	An editor describes the origins of his/her statements.	editor	Specify a statement by statements affirming this relation.	[Requirement 73]	Statements can be interlinked with information about their origin.
Scenarios concerning the usability of resources					
Scenario 43	A user found information about an item and wants to know how he/she may get access to the item.	user	Browse from item description to the description of the terms of use	[Requirement 59]	Item descriptions must provide machine readable information about the accessibility and usability of items.
			Browse from item description to the description of the repository/place	[Requirement 77]	Item descriptions must interlink with information about the current location/place of
				[Requirement 83]	Holdings must be identified by identifiers that are unique, machine readable and persistent.
				[Requirement 85]	Holding descriptions must include 1-n place appellations
Scenario 44	A user wants to contact the agents who may help him/her to use items of a collection.	user	Browse from item description to the description of the repository/place	[Requirement 77]	Item descriptions must interlink with information about the current location/place of the item.
			Find contact information at the description of the repository/place	[Requirement 58]	Holding descriptions must provide contact information.
Scenario 45	A user is searching for an item in an institution	user	Find a resource near to your	[Requirement	Item descriptions must interlink with

Schritt 4: Identifikation von Klassen und Beziehungen zwischen den Klassen



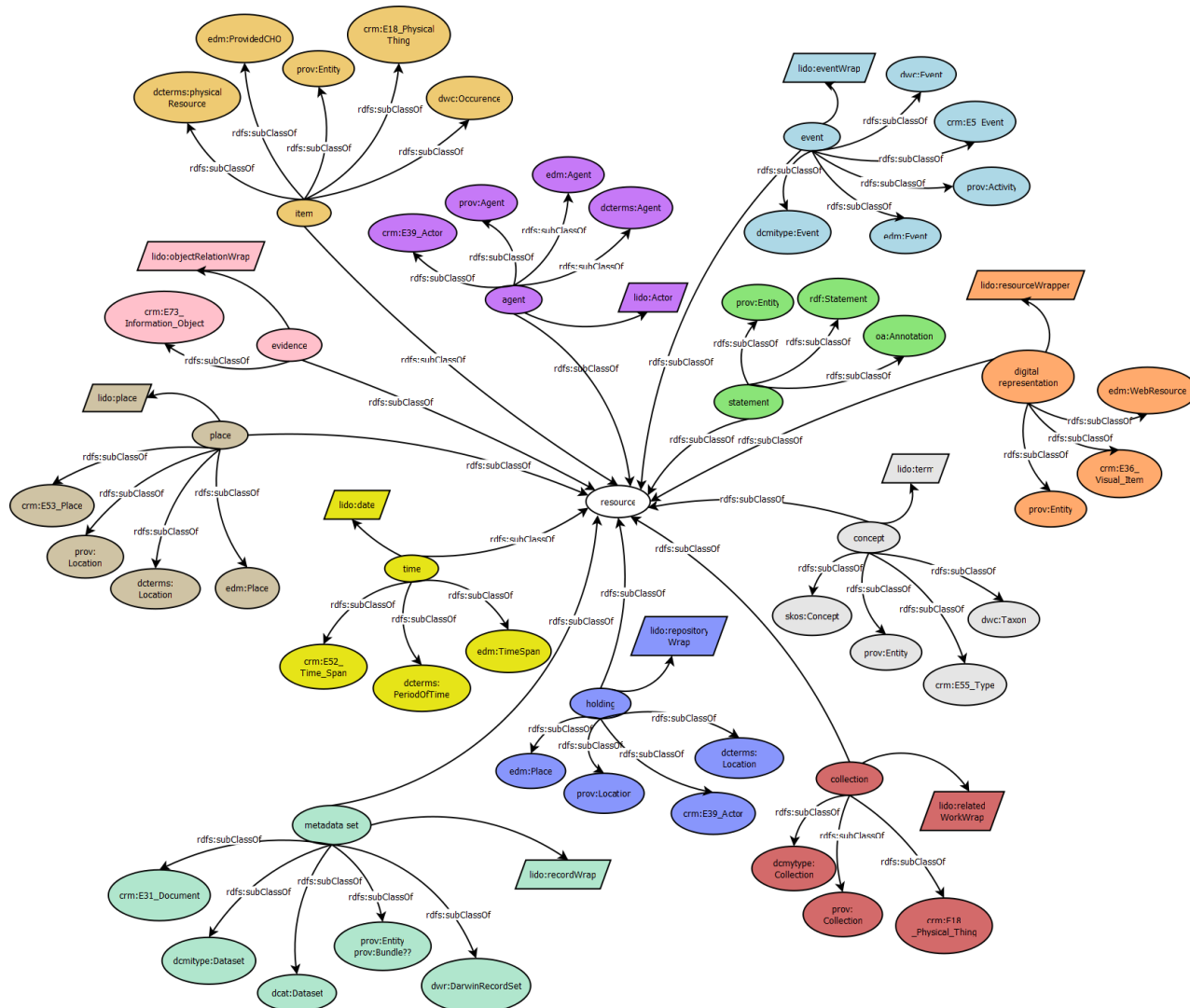
- Liste der Klassen
 - resource
 - metadata set
 - item
 - evidence
 - event
 - time
 - agent
 - place
 - **digital representation**
 - collection
 - statement
 - holding
 - concept

Abgleich der Klassen und Requirements mit einschlägigen Standards

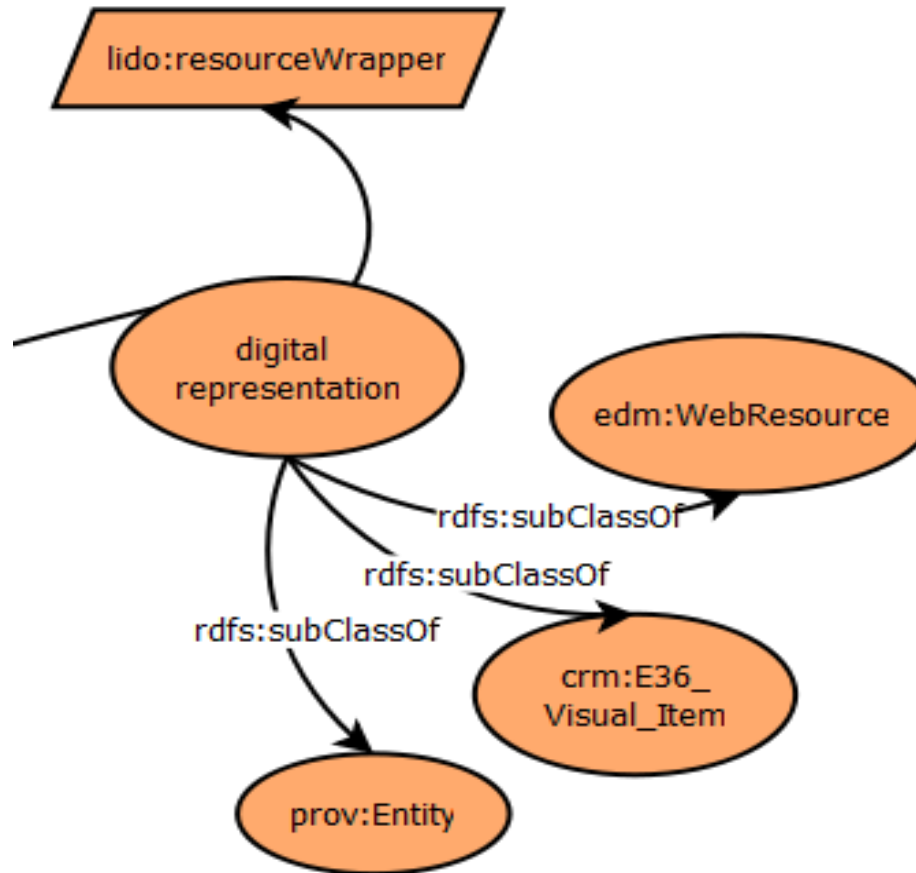


Digital Representation	A digital resource depicting an item	Entity	E36 Visual Item	This class comprises the intellectual or conceptual aspects of recognizable marks and images.		✓
[Requirement 43]	A digital representation/copy must enable the identification of the item it represents.			Out of scope.		✗
[Requirement 45]	The nature of a digital representation must be described using a controlled value.	Property	P2 has type	This property allows sub typing of CRM entities – a form of specialization – through the use of a terminological hierarchy, or thesaurus.	crm:E36_Visual_Item crm:P2_has_Type crm:E55_Type	✓
[Requirement 60]	Descriptions of digital representations must provide information about the reusability of the digital representation (e.g. license).	Property	P104 is subject to	This property links a particular E72 Legal Object to the instances of E30 Right to which it is subject.	crm:E36_Visual_Item crm:P104_is_subject_to crm:E30_Rights crm:E30_Rights crm:P75i_is_posessed_by crm:E39_Actor	✓
[Requirement 61]	Descriptions of a digital representation must provide contact information related to the usability of a digital copy.	Property	P75i is possessed by / P76 has contact point	This property identifies former or current instances of E30 Rights held by an E39 Actor. / This property identifies an E51 Contact Point of any type that provides access to an E39 Actor by any communication method, such as e-mail or fax.	crm:39_Actor crm:P76_has_contact_point crm:E51_Contact_Point	✓
[Requirement 82]	Digital representations must be identified by identifiers that are unique, machine readable and persistent.	Property	P1 is identified by	The property describes the naming or identification of any real world item by a name or any other identifier.	crm:E36_Visual_Item crm:P1_is_identified_by crm:E42_Identifier .	✓
Collection	An aggregation of items:	Entity	E18 Physical Thing	This class comprises all persistent physical items with a relatively stable form, man-made or natural		✓

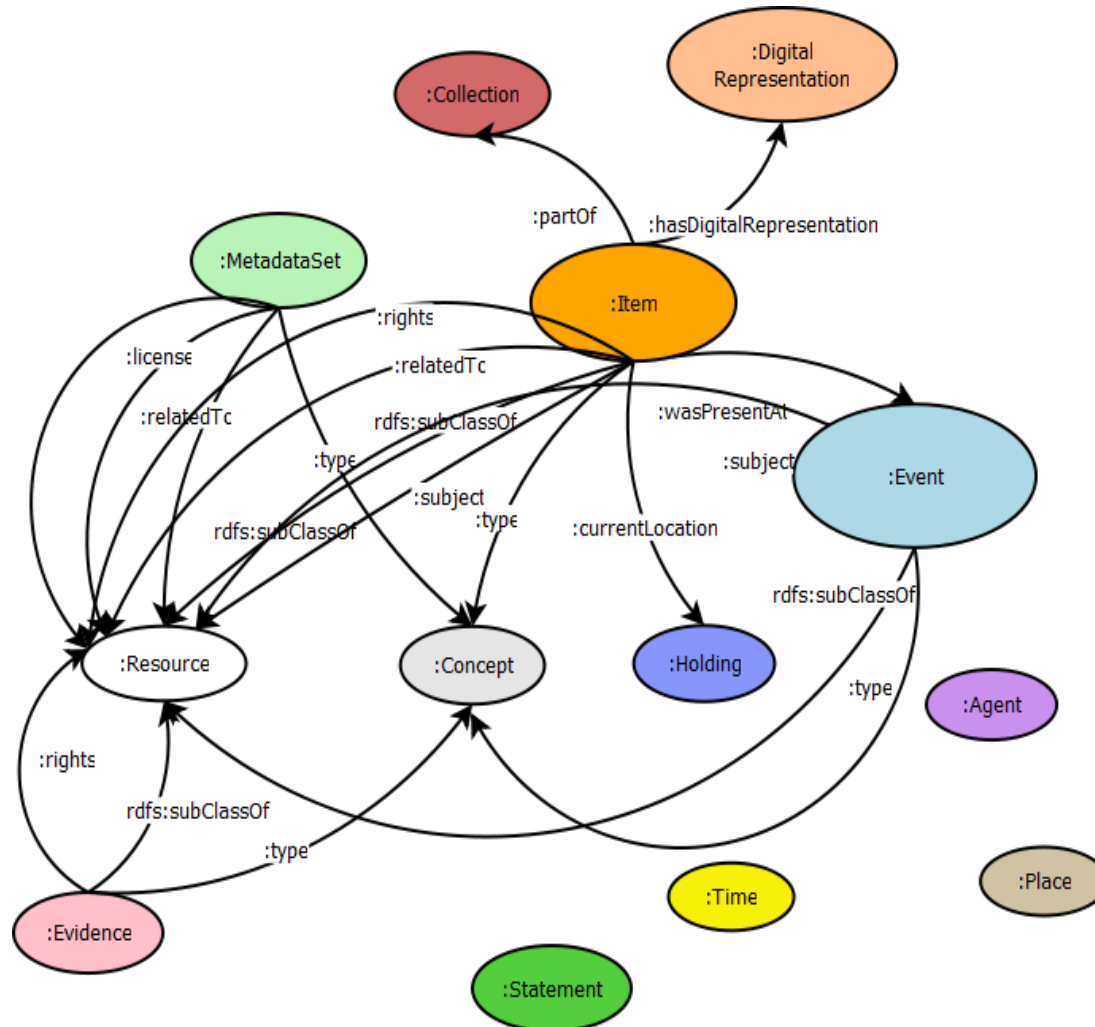
Mapping der Klassen (Entitäten) auf Standards



Beispiel: Mapping der Entität *digital representation*



Entity-Relationship-Model (in Arbeit)



Schritt 5: Identifikation von Properties (in Arbeit)



- Auf der Grundlage der Requirements wurden zunächst standardunabhängige Properties entwickelt:

Digital Representation

:DigitalRepresentation

Type of term: rdfs:Class

Definition: A digital resource depicting an item.

Properties used with :DigitalRepresentation

:identifier/about

Type of Term: rdfs:Property

Definition: relates to a unique, machine readable and persistent identifier identifying this metadata set.

Domain: Resource

Range: Literal

Occurrence: ???

Beispiel XML und RDF XML und RDF Turtle

Subproperty of:

:type

Type of term: rdfs:Property

Definition: relates to a controlled value describing the nature of the resource.

Domain: Resource

Range: skos:Concept

Occurrence: 1-n

:license

Type of term: rdfs:Property

Definition: Information about the reusability of the MetadataSet.

Domain: MetadataSet

Range: LicenseDocument

Occurrence: 1

Note: Use Creative Commons Licenses (<https://creativecommons.org/>)

- Diese Properties werden im nächsten Schritt in Form von Standard-Ontologien umgesetzt
- Momentan: Test für die Umsetzung in CIDOC-CRM



:DigitalRepresentation			<u>crm:E36_Visual_Item</u>	rdf:about	
:identifier	URI des Digitalisats	crm:E36_Visual_Item	crm:P1_is_identified_by	crm:E42_Identifier	
:type	Art des Digitalisats	crm:E36_Visual_Item	crm:P2_has_type	crm:E55_Type	
:license	Rechte am Digitalisat	crm:E36_Visual_Item	crm:P104_is_subject_to	crm:E30_Right	
:relatedTo		crm:E36_Visual_Item	crm:P130_shows_features_of	crm:E70_Thing	
		crm:E36_Visual_Item	crm:P138_represents	crm:E1_CRM_Entity	
			crm:P67.1_has_type	crm:E55_Type	
			crm:P138.1_mode_of_representation	crm:E55_Type	
:rightsHolder		crm:E36_Visual_Item	crm:P105_right_held_by	crm:E39_Actor	
:subject		crm:E36_Visual_Item	crm:P129_is_about	crm:E1_CRM_Entity	
			crm:P67.1_has_type	crm:E55_Type	
:depicts		crm:E36_Visual_Item	foaf:depicts	crm:E1_CRM_Entity	
:Collection			<u>crm:E78_Collection</u>	rdf:about	
:identifier	ID der Sammlung	crm:E78_Collection	crm:P1_is_identified_by	crm:E42_Identifier	
:name	Name der Sammlung	crm:E78_Collection	crm:P1_is_identified_by	crm:E41_Appellation	
:type	Art der Sammlung	crm:E78_Collection	crm:P2_has_type	crm:E55_Type	
		crm:F78_Collection	crm:P137_exemplifies	crm:F55_Type	



Auswahl geeigneter Standards

:MetadataSet			<u>crm:E31_Document</u>	rdf:about	
:identifier	ID des Datensatzes	crm:E31_Document	dc:identifier		
:type	Art der Ressource	crm:E31_Document	dc:type		
:relatedToAgent	Creator	crm:E31_Document	dc:creator		
:dateCreated	Entstehungsdatum	crm:E31_Document	dcterms:created		
:relatedToAgent	Bearbeiter	crm:E31_Document	dc:contributor		
:dateModified	Bearbeitungsdatum	crm:E31_Document	dcterms:modified		
:relatedToAgent	Datengeber	crm:E31_Document	edm:provider		edm:Agent
:license	CC-Lizenz	crm:E31_Document	dcterms:license		dcterms:LicenseDocument
:conformsTo	Link zum Standard, Schema oder AP	crm:E31_Document	dcterms:conformsTo		dcterms:Standard
:relatedTo	Link zum beschriebenen Objekt	crm:E31_Document	crm:P129_is_about		crm:E70_Thing
			(crm:P67.1_has_type		crm:E55_Type)
:Item			<u>crm:E70_Thing</u>	rdf:about	
:identifier	ID des Objekts	crm:E70_Thing	crm:P1_is_identified_by		crm:E42_Identifier
:name	Name/Bezeichnung des Objekts	crm:E70_Thing	crm:P1_is_identified_by		crm:E41_Appellation
:type	Art des Objekts	crm:E70_Thing	crm:P2_has_type		crm:E55_Type
		crm:E70_Thing	crm:P137_exemplifies		crm:E55_Type



Schritt 6: Erstellen der Anwendungsprofile

- Dies ist der nächste Schritt ...
- Nachnutzung dessen, was schon da ist
 - Fokus liegt auf Provenienzinformation
 - W3C PROV
 - Berücksichtigung anderer Standards (EDM, CIDOC-CRM, ...)
 - Nutzung domainspezifischer Standards (LIDO, MODS, DC, ...)
- Einbindung der Evidenzen
- Einbinden der Anwendungsprofile in das Entity-Relationship-Modell

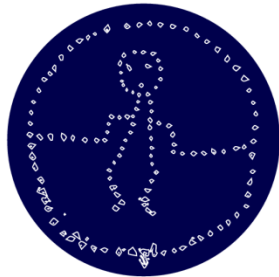


Schritt 7: Test

- Modelltest in verschiedenen Systemen:
 - WissKI → CIDOC-CRM
 - Semantic Media Wiki
 - OntoWiki
 - ...

SUB

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



Danke für Ihre Aufmerksamkeit!

zum Projekt: <http://asch.wiki.gwdg.de>

